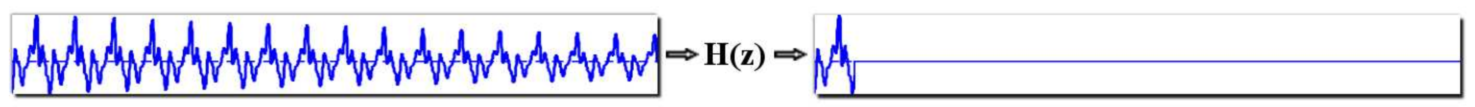
Cascaded Long Term Prediction for Coding Polyphonic Audio Signals

Tejaswi Nanjundaswamy and Kenneth Rose

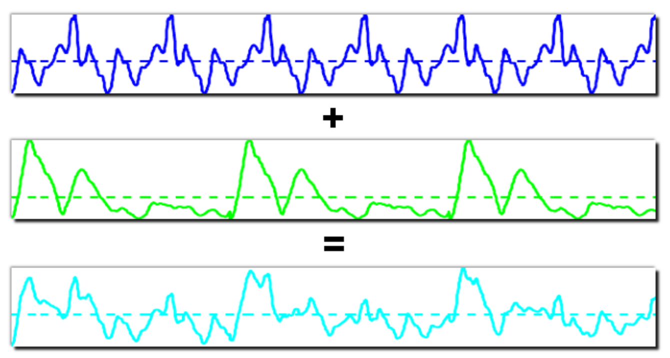
Signal Compression Lab, Department of ECE, University of California Santa Barbara

Motivation

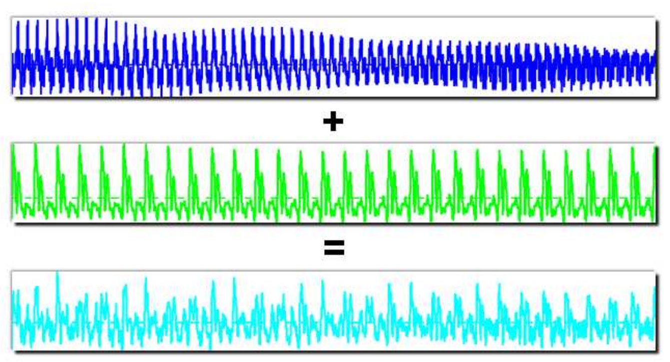
- Most audio signals contain periodic components
- Predicting these components efficiently is crucial for coding
- For monophonic signals prediction is a solved problem and long term prediction (LTP) is a well known efficient solution
- LTP filter $H(z) = 1 - \alpha z^{-N}$ identifies a similar previous segment and scales it as prediction for current segment



- LTP clearly effective for monophonic files, but most signals are polyphonic



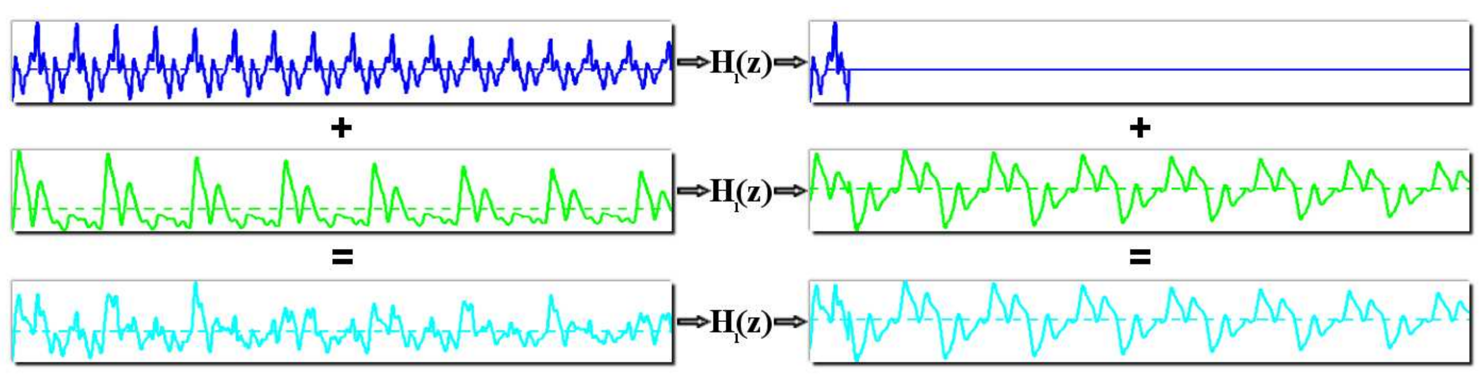
- In principle this mixture is periodic, but the new period is too long (at LCM of all periods) and real audio signal rarely remains stationary for such durations



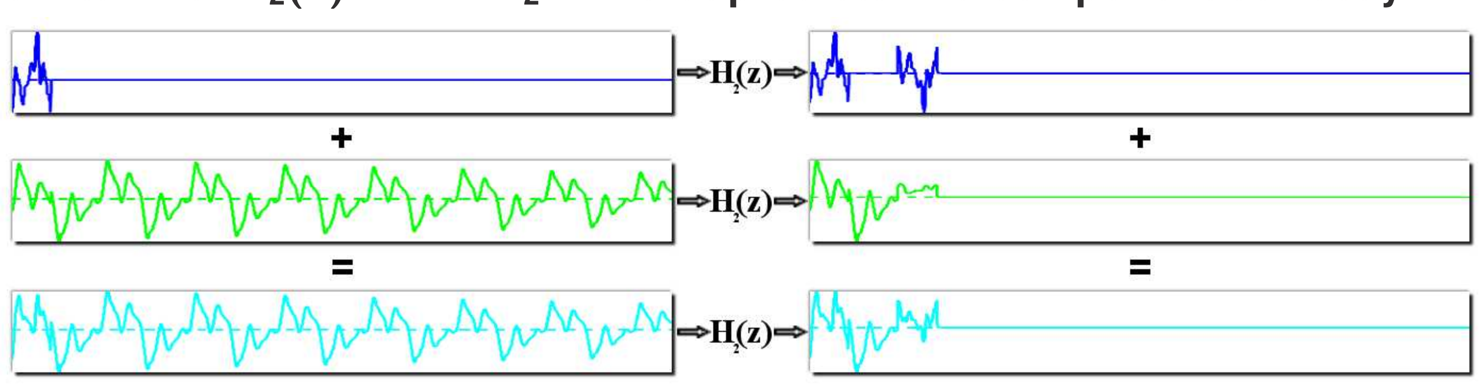
- LTP is suboptimal here as no waveform repetition

How to predict a polyphonic file?

- Let's start with filter $H_1(z) = 1 - \alpha_1 z^{-N_1}$ that predicts 1st component correctly
- Mixture residue got as sum of every component's residue



- Note that periodicity retained in 2nd component even after filtering
- So filter with $H_2(z) = 1 - \alpha_2 z^{-N_2}$ that predicts 2nd component correctly



Cascaded long term prediction

- Clearly encoding this residue results in compression gains
- Thus the cascaded long term prediction (CLTP) filter forms the basis of this proposal

$$H_c(z) = \prod_{i=0}^{P-1}(1 - \alpha_i z^{-N_i} - \beta_i z^{-N_i+1})$$

- Note that each filter in the cascade is 2nd order so that α and β can now capture non-integral pitch periods as well

CLTP parameter estimation

- As noted before periodicity of a component is not altered by a filter not designed for it
- Thus parameters of jth filter are estimated in the residue $\hat{x}_j[m]$ after filtering with all the other filters

$$H_j(z) = \prod_{\forall i, i \neq j}(1 - \alpha_i z^{-N_i} - \beta_i z^{-N_i+1})$$

- Estimating parameters of one filter $1 - \alpha_j z^{-N_j} - \beta_j z^{-N_j+1}$ is simply the well known LTP problem
- Given N, the $\alpha_{(j,N)}$ and $\beta_{(j,N)}$ is based on the correlation and given as

$$\begin{bmatrix} \alpha_{(j,N)} \\ \beta_{(j,N)} \end{bmatrix} = \begin{bmatrix} r_{(N,N)} & r_{(N-1,N)} \\ r_{(N-1,N)} & r_{(N-1,N-1)} \end{bmatrix}^{-1} \begin{bmatrix} r_{(0,N)} \\ r_{(0,N-1)} \end{bmatrix}$$

where the correlation values $r_{(k,l)}$ are

$$r_{(k,l)} = \sum \hat{x}_j[m-k]\hat{x}_j[m-l]$$

- The best N_j is found as the one which minimizes the mean squared error

$$N_j = \underset{N \in [N_{\min}, N_{\max}]}{\arg\min} \sum \begin{pmatrix} \hat{x}_j[m] - \alpha_{(j,N)}\hat{x}_j[m-N] \\ - \beta_{(j,N)}\hat{x}_j[m-N+1] \end{pmatrix}^2$$

- This process is repeated in a loop until convergence
- Convergence is guaranteed as at each step, overall prediction error is monotone non-increasing

Integration with Bluetooth Sub-band Codec (SBC)

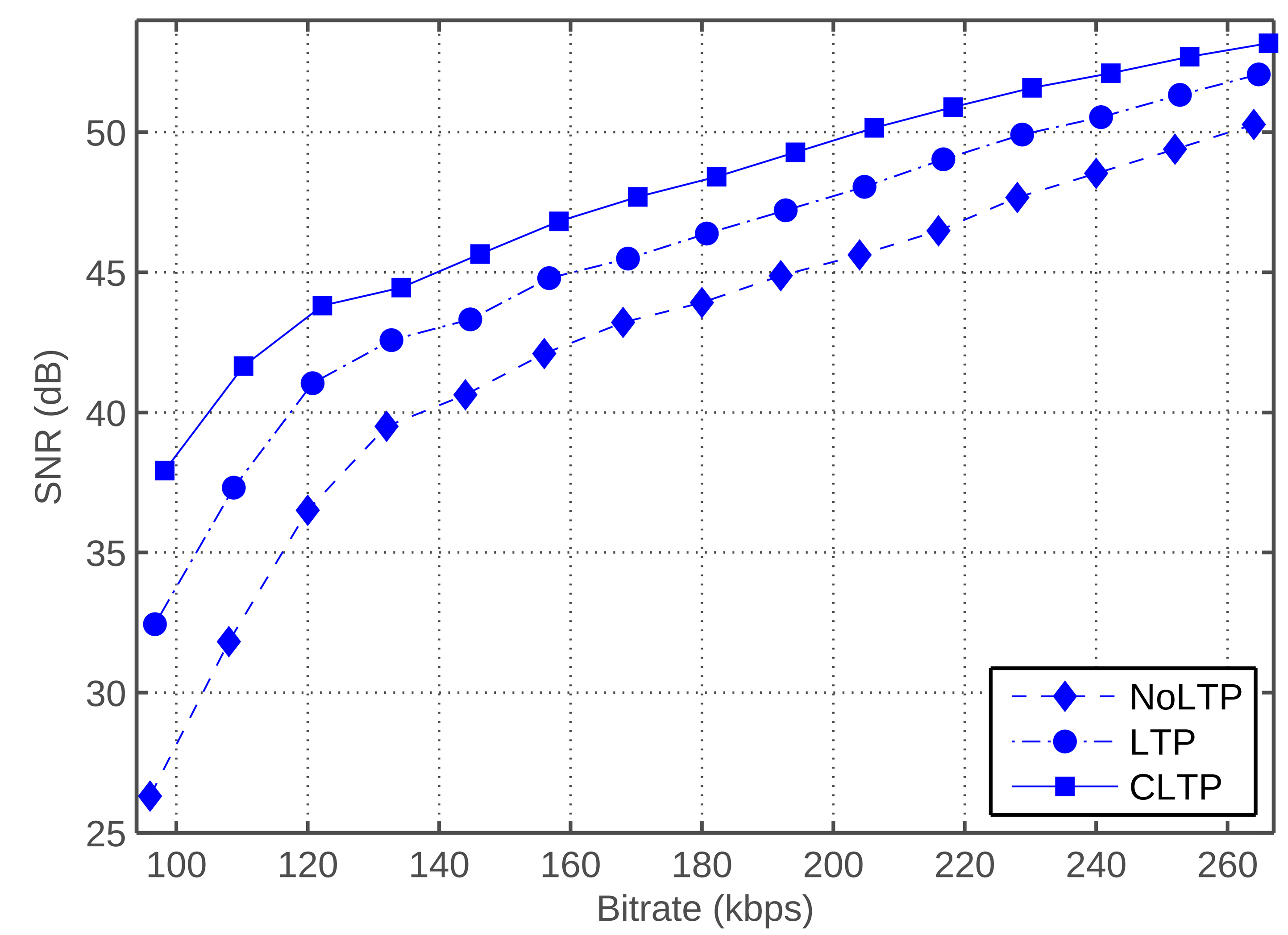
- SBC is an ultra-low-delay coder which analyzes signal into sub-bands and adaptively quantizes sub-band samples in small blocks
- Clearly SBC's capability to exploit redundancies is limited to small block lengths
- Thus CLTP integrated to improve compression efficiency by providing effective inter-block prediction
- CLTP employed only in the first sub-band to ensure simplicity, while selectively predicting the critical low frequencies effectively
- The CLTP filter parameters are estimated backward adaptively
  - Assumes signal to be locally stationary
  - Reduces side-information rate

Evaluations

- Comparison: Reference SBC with no prediction, SBC with a single LTP filter, SBC with the proposed CLTP
- Testing data-set: 7 audio files, 4s each, mono, 44.1/48 kHz
- Objective evaluation results with prediction gains and reconstruction gains in dB

	Prediction gains		Reconstruction gains	
Filename	LTP	CLTP	LTP	CLTP
Piano	5.8	15.0 (+9.2)	3.2	6.9 (+3.7)
Guitar	9.5	15.9 (+6.4)	5.0	7.9 (+2.9)
Harp	6.5	14.4 (+7.9)	5.8	12.6 (+6.8)
Bells	6.0	16.7 (+10.7)	5.4	13.9 (+8.5)
Mfv	11.6	19.0 (+7.4)	11.5	16.8 (+5.3)
Mozart	7.9	15.4 (+7.5)	6.3	11.5 (+5.2)
Quartet	3.0	7.3 (+4.3)	2.3	5.7 (+3.4)
Average	7.2	14.8 (+7.6)	5.6	10.8 (+5.2)

- Operational rate-distortion (RD) plots of SNR versus bit-rate



Conclusions

- Existing periodic component predicting technique of LTP sub-optimal for polyphonic signals
- Cascading LTP filters to optimally predict polyphonic signals proposed
- An effective recursive technique for estimation of filter parameters proposed
- Evaluations within the Bluetooth SBC substantiates the effectiveness of the proposed approach
- Future directions includes adapting CLTP to perceptual coders like MPEG AAC

http://www.scl.ece.ucsb.edu/ {tejaswi, rose}@ece.ucsb.edu